AFRL-AFOSR-VA-TR-2016-0009

Controlled Visual Sensing and Exploration

Stefano Soatto
UNIVERSITY OF CALIFORNIA LOS ANGELES

09/16/2015
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
06-12-2015

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
01-07-2012 to 30-06-2015

4. TITLE AND SUBTITLE
Controlled Visual Sensing and Exploration

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0364

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Stefano Soatto

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF CALIFORNIA LOS ANGELES
11000 KINROSS AVE STE 102
LOS ANGELES, CA 90095-0001 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project developed analytical and computational tools to design integrated sensor-control systems, where the controller is part of the sensor and designed so as to maximize task-specic information. Within this broad umbrella, we have focused on visual sensors (EO imagery), inertial sensors (accelerometers and gyrometers) and ranging sensors (structured light), and their integration in support of mobility task (exploration) and decisions (detection, localization, recognition, categorization of objects and scenes). The task informs what part of the data-formation process is a nuisance, i.e. it is irrelevant to the task but nevertheless aects the data. Obviously, the resulting sensor-control system depends on the data and it depends on the task. We have focused on tasks that require invariance or co-variance to illumination and to vantage point. Then the control reduces to mobility of the sensor platform, so as to overcome occlusion or scaling limitation in the passive version of the sensor. Therefore, the actuation, control, and sensing systems are collectively considered an active sensor, and algorithms for inference, planning and control can be co-designed so as to achieve maximum uncertainty reduction in the task, or maximum actionable information [11].

15. SUBJECT TERMS
Visual Sensing

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UU		Stefano Soatto
					19b. TELEPHONE NUMBER *(Include area code)* 310-825-4840

CONTROLLED VISUAL SENSING AND EXPLORATION AFOSR FA9550-12-1-0364

Stefano Soatto
UCLA Vision Lab
University of California, Los Angeles

Final Report
May 1, 2015

Abstract

1 Summary of Research Achievements

This project developed analytical and computational tools to design integrated sensor-control systems, where the controller is part of the sensor and designed so as to maximize task-specific information. Within this broad umbrella, we have focused on visual sensors (EO imagery), inertial sensors (accelerometers and gyrometers) and ranging sensors (structured light), and their integration in support of mobility task (exploration) and decisions (detection, localization, recognition, categorization of objects and scenes). The task informs what part of the data-formation process is a *nuisance*, i.e. it is irrelevant to the task but nevertheless affects the data. Obviously, the resulting sensor-control system depends on the data and it depends on the task. We have focused on tasks that require invariance or co-variance to illumination and to vantage point. Then the control reduces to mobility of the sensor platform, so as to overcome occlusion or scaling limitation in the passive version of the sensor. Therefore, the actuation, control, and sensing systems are collectively considered an active sensor, and algorithms for inference, planning and control can be co-designed so as to achieve maximum uncertainty reduction in the task, or maximum *actionable information* [11].

1.1 Modeling the agent

The first ingredient to establish an active remote sensor is the ability to move, which requires the ability to localize the sensor platform, or *agent*, relative to the surrounding environment.

Thus, inferring (causally, and in real-time) a representation of the environment and the agent's location (position and orientation relative to it) is a key enabler and a fundamental and classical problem in a number of fields. It is well known [10], for instance, that pose (a trajectory in the Lie group $SE(3)$) can be inferred *up to a spatial similarity transformation* by a monocular EO sensor, under suitable conditions. However, knowledge of scale is essential for interaction, so EO-only approaches are not suitable for active sensing control. Traditionally, inertial navigation provides scaled-estimates of pose, but without reference to the surrounding environment and based on a doubly-integrated non-observable model that yields diverging error dynamics. Thus our first accomplishment is to study the integration of visual and inertial sensing, and the development of what we believe to be the most advanced platform for visual-inertial fusion: [9] shows feasibility, [15] shows flexibility, [14] shows robustness. Part of this work will be presented at the next ICRA (International Conference on Robotics and Automation) where the first is short-listed for Best Conference Paper. The critical element of this work is its focus on *robustness*, for what we have shown [12] is that most of visual data is useless for most tasks, and therefore one can expect – as indeed happens – that most of the data consists of *outlier measurements*. Unlike traditional filtering stemming from the field of robust statistics, in the scenario of interest it is typical to have a majority of *outlier* measurements. It has been necessary, therefore, to revisit classical robust filtering to handle these scenarios, which has been accomplished during the project. Specific accomplishments in this portion of the project includes:

- We have shown that commonly used models for visual inertial fusion are *not* observable/identifiable. While they would be identifiable if accel and gyro bias rates were *known* or constant, in general they are not. This (negative) result undermines much of the prior analysis of observability and identifiability of visual-inertial sensor fusion.

- While not observable, we have shown that the indistinguishable set of state trajectories is *bounded*, and we have computed it explicitly as a function of sensor characteristics and motion statistics.

- We have used the analysis to derive a model for a nonlinear filter that is then used to converge to a state in a set, and we have bounds for said set.

- We have designed an outlier rejection algorithm based on a finite whiteness test (Box-Ljung) computed on a temporal sliding window, and a causal smoothing scheme to support its computation, which is shown to approximate the optimal (Neymann-Pearson) discriminant.

- We have demonstrated the system live at CVPR, benchmarked against Google Tango – a project that benefits from corporate backing and over 20 engineers working full-time on it for over 2 years – outperforming it despite a single graduate student effort.

- The paper is a finalist for Best Conference Paper at the next ICRA.

1.2 Modeling the scene

Localization is only the first step to enable spatial interaction and decision tasks concerning the scene. The representation of the underlying environment sufficient to support localization is typically a sparse point cloud. This is clearly insufficient for most other tasks that require at least the *topology* of the scene to determine what surfaces or "objects" are neighbors. For instance, for navigation, it is vital to know whether the empty space between two points is occupied by their supporting surfaces, or whether it is empty space, as in the latter case it is traversable, in the former it is not.

To this end, we have developed methods for topology estimation and regularization, as well as coupling between location estimation and coarse geometry: [4] couples the two, [3] uses technique for range imagery, [7] develops robust methods for densification and reconstruction, [8] regularizes with the structure tensor. Furthermore, [5] develops first second-order method for geometric inverse problems.

As part of this effort, we have performed **analysis and design of co-variant detectors and their associated invariant descriptors** (low-level, local descriptors [6]) and **dynamic scene analysis** [13], leveraging work on **occlusion detection** [1, 2]. Specific achievements include:

- We have shown that surface topology and geometry can be computed without *minimal surface bias*, yielding water-tight surfaces and accurate volume measurements. These have been used by neuroscientists to study the perceptual bias in the relation between size and weight of objects.

- We have developed novel regularizers that respect the surface geometry without the need to know their topology, exploiting instead the (trivial) image topology. This means that we can run dense reconstruction in real time.

- We have developed the *first* second-order (Newton-like) optimization scheme on geometric shape spaces.

- We have leveraged on prior work on occlusion detection to develop scene partition schemes that can account for individual objects' motion and relative occlusion, while maintaining persistent tracking.

1.3 Controlling the sensor

In [16] we have shown how the controller can be part of the sensor and collectively make a system "the best sensor it can be", in the sense of controlling the data acquisition process so as to minimize task uncertainty. While in the early phases of the project this construction was restricted to cartoon two-dimensional objects, as the ensuing optimization problem becomes quickly intractable and in any case beyond real-time low-latency implementation suitable to closing the loop, during the latter phase of the project we have developed efficient computational approximations based on extended using Poisson sampling that have enabled

us not only to extend it to 3D but also to allow non-compact domains, essential in remote sensing such as vision and ranging. Specific achievements include

- We have extended the optimal exploratory control work, previously developed, to non-compact domains. In order to compute uncertainty reduction, one needs to compute the "probability of visibility," which is the probability of sensing portions of the scene that are occluded. This requires a prior. If the domain is not compact, the uncertainty is infinite, and therefore the uncertainty reduction (information) is not defined. We have developed a method based on Poisson Sampling that makes this sound mathematically, and efficient to compute, with Poisson-Voronoi partitions.

- We have tested Poisson-Voronoi based planning and (pseudo-)optimal control in simulated environments in 2D and 3D

While all the milestones foreseen in the original proposal have been met, new paths forward have opened during the investigation. Specifically, now that the formalization of the problem of maximizing "actionable information" has been done, there remains the need to derive tractable approximations that come with some kind of performance guarantee. The work on Poisson-Voronoi sampling is one such example, but much more work is needed to extend this work to more complex tasks, and to higher level of abstractions of the scene, where the interaction and control is not only based on geometry and topology, but on the *semantics* of the scene, that is its partition into objects and the description of their relations.

This is part of future work that we intend to commence now.

Acknowledgment/Disclaimer

This work was sponsored in part by the Air Force Office of Scientific Research, USAF, under grant number FA9550-09-1-0427. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, or the U.S. Government.

References

[1] A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *Intl. J. of Comp. Vision*, 97(3):322–338, 2012.

[2] A. Ayvaci and S. Soatto. Detachable object detection: Segmentation and depth ordering from short-baseline video. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 34(10):1942–1951, 2012.

[3] J. Balzer, M. Peters, and S. Soatto. Volumetric reconstruction applied to perceptual studies of size and weight. In *Proc. of WACV*, March 2014 2014.

[4] J. Balzer and S. Soatto. Clam: Coupled localization and mapping with efficient outlier handling. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2013.

[5] J. Balzer and S. Soatto. Second-order shape optimization for geometric inverse problems in vision. *Proc. of CVPR*, Nov. 8, 2013.

[6] J. Dong and S. Soatto. Domain size pooling in local descriptors: Dsp-sift. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, 2015.

[7] V. Estellers, M. Scott, K. Tew, and S. Soatto. Robust poisson surface reconstruction. In *Scale-Space Variational Methods (SSVM)*, 2015.

[8] V. Estellers, S. Soatto, and X. Bresson. Adaptive regularization with the structure tensor. *IEEE Trans. on Image Processing*, 2015.

[9] J. Hernandez, K. Tsotsos, and S. Soatto. Observability, identifiability and sensitivity of vision-assisted inertial navigation. *Proc. of the Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[10] S. Soatto. Observability/identifiability of rigid motion under perspective projection. In *33 IEEE Conf. on Decision and Control*, pages 3235–3240, Dec. 1994.

[11] S. Soatto. *Steps Toward a Theory of Visual Information.* http://arxiv.org/abs/1110.2053, Technical Report UCLA-CSD100028, September 13, 2010 2010.

[12] G. Sundaramoorthi, P. Petersen, V. S. Varadarajan, and S. Soatto. On the set of images modulo viewpoint and contrast changes. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, June 2009.

[13] B. Taylor, V. Karasev, and S. Soatto. Causal video object segmentation from persistence of occusion. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, 2015.

[14] K. Tsotsos, A. Chiuso, and S. Soatto. Robust filtering for visual inertial sensor fusion. *Proc. of the Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[15] K. Tsotsos, A. Pretto, and S. Soatto. Visual-inertial ego-motion estimation for humanoid platforms. In *Proc. IEEE Intl. Conf. Humanoid Robots*, Dec. 2012.

[16] L. Valente, Y.-H. Tsai, and S. Soatto. Information-seeking control under visibility-based uncertainty. *J. of Math. Im. and Vis.*, 2013.

Personnel Supported During Duration of Grant

Stefano Soatto (PI)
Jingming Dong (Graduate Student)
Konstantine Tsotsos (Graduate Student)
Zhaoking Bu (Graduate Student)
Jonathan Baler (Postdoc)
Vasiliy Karasev (Graduate Student)

Publications (see section "References" above)

Honors & Awards Received

IEEE Fellow, 2013; **International Computer Vision Summer School,** Invited Speaker, Scicli, July 2012, 2013, 2014, 2015; **International Conference on Computer Vision,** Program Co-Chair, Venice, Italy, Oct. 2017; **International Journal of Computer Vision**, Associate Editor; **Journal of Mathematical Imaging and Vision**, Associate Editor; **Foundations and Trends in Graphics and Vision**, Associate Editor; **SIAM Journal of Imaging Science**, Associate Editor.

AFRL Point of Contact

Dr. Fariba Fahroo, Program Manager, Computational Mathematics, AFOSR/NL, 875 North Randolph Street, Suite 325, Room 3112, Arlington, VA 22203, (703) 696-8429, Fax (703) 696-8450, DSN 426-8429, fariba.fahroo@afosr.af.mil

Transitions: Software developed during the project was integrated within the open-source repository "VlFeat" (www.vlfeat.org). Software for visual-inertial navigation has been distributed to numerous academic and government laboratories, including ONR China Lake.

New Discoveries: Most publications have been supplemented by open-source code and distributed freely for non-commercial purposes. No patented disclosures were submitted as part of the research conducted in this project.

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

soatto@ucla.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

3108254840

Organization / Institution name

UCLA

Grant/Contract Title
The full title of the funded effort.

Controlled VIsual Sensing and Exploration

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0364

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Stefano Soatto

Program Manager
The AFOSR Program Manager currently assigned to the award

Fariba Fahroo

Reporting Period Start Date

05/01/2012

Reporting Period End Date

05/01/2015

Abstract

This project developed analytical and computational tools to design integrated sensor-control systems, where the controller is part of the sensor and designed so as to maximize task-specific information. Within this broad umbrella, we have focused on visual sensors (EO imagery), inertial sensors (accelerometers and gyrometers) and ranging sensors (structured light), and their integration in support of mobility task (exploration) and decisions (detection, localiza- tion, recognition, categorization of objects and scenes). The task informs what part of the data-formation process is a nuisance, i.e. it is irrelevant to the task but nevertheless affects the data. Obviously, the resulting sensor-control system depends on the data and it depends on the task. We have focused on tasks that require invariance or co-variance to illumination and to vantage point. Then the control reduces to mobility of the sensor platform, so as to overcome occlusion or scaling limitation in the passive version of the sensor. Therefore, the actuation, control, and sensing systems are collectively considered an active sensor, and algorithms for inference, planning and control can be co-designed so as to achieve maximum uncertainty reduction in the task, or maximum actionable information [11].

Distribution Statement
This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

AFD-070820-035.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

FinalReport2015Soatto.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

see uploaded PDF

Changes in research objectives (if any):

n/a

Change in AFOSR Program Manager, if any:

n/a

Extensions granted or milestones slipped, if any:

n/a

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Sep 03, 2015 21:36:15 Success: Email Sent to: soatto@ucla.edu